

Sierra Center of Excellence Overview

DOE Centers of Excellence Performance Portability Meeting

Glendale, AZ

April 19, 2016

Rob Neely

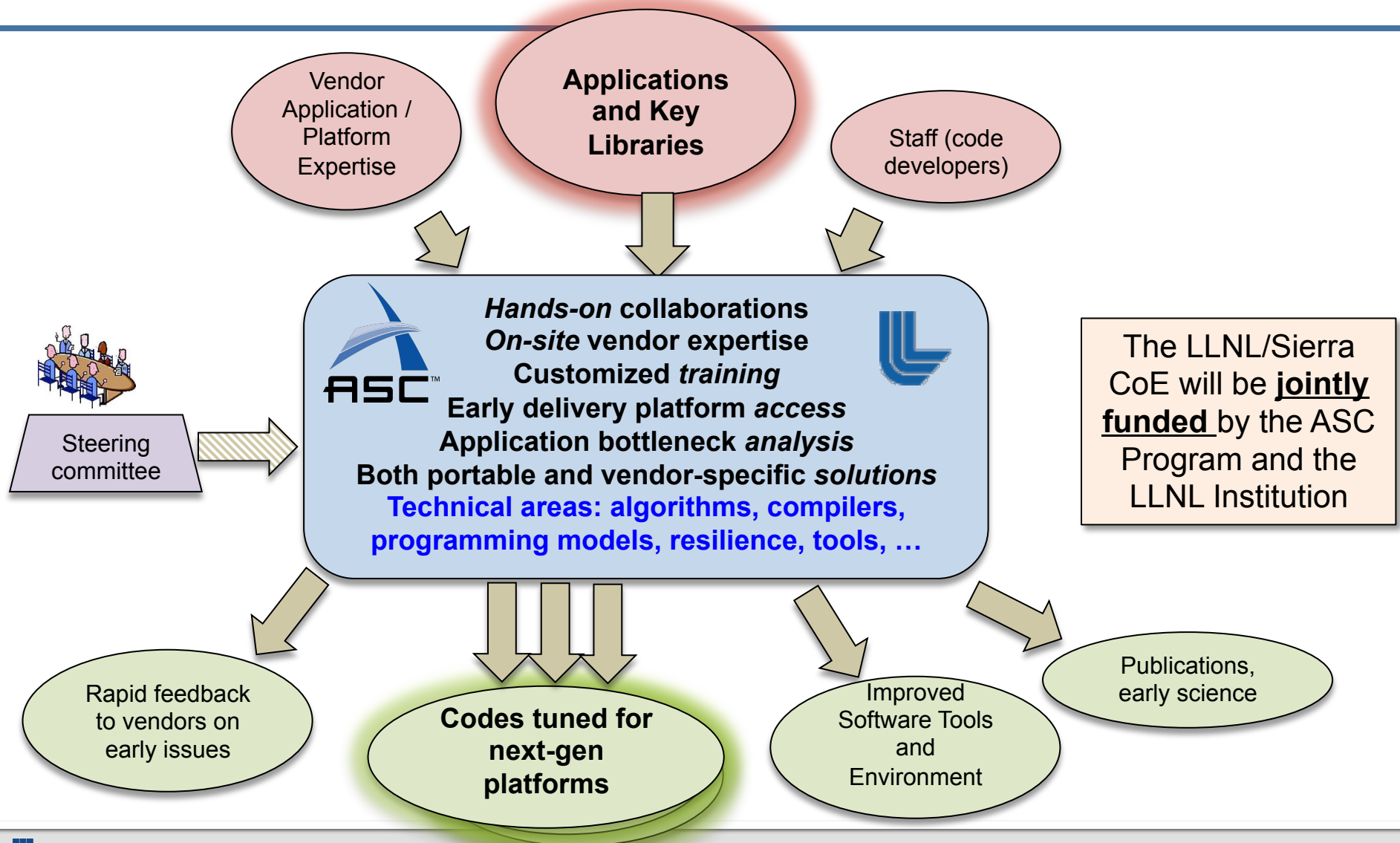


LLNL-PRES-690092

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

 **Lawrence Livermore
National Laboratory**

Application enablement through the Center of Excellence – the original (and enduring) concept



ASC applications receiving COE support determined by mission-need

- ASC labs have a set of mission-critical ASC applications
 - Proxy apps (completely open, useful for initial studies)
 - Unclassified (but restricted access) applications
 - Classified applications
- ASC Apps targeting our use of Sierra were typically developed over the past 10-20 years
 - ATDM “from scratch” applications also to be included
 - Targeting first production use during lifetime of *Sierra*
- Time on the machines is granted through a tri-lab proposal process
 - CCC’s – Capability Computing Campaigns
- Application preparation for new platforms is a standard part of each teams mandate
 - Multi-disciplinary teams (CS + physics + engineering expertise)
 - Progress/reviews performed through ASC milestones (L1/L2/L3)



COE vendor participation amplifies and accelerates team efforts



Starting in FY16, we expanded the *Sierra* COE to include non-ASC LLNL “institutional” applications

- LLNL has long provided institutional hardware for general unclassified use
 - Smaller versions of large classified ASC investments
 - E.g. *Vulcan* – 5PF BG/Q system (25% of *Sequoia*). Other large linux clusters
 - Planned installation of “mini-Sierra” in 2018
- Needed: similar effort to prepare institutional applications that don’t receive ASC support
- ICOE (Institutional COE)
 - Mixture of internal and vendor support (code teams + AAPS team)
 - Projects selected through internal process
 - Increasing focus on data-centric computing needs
 - Managed by Bert Still



The ICOE will ensure LLNL-developed strategic codes can effectively use CORAL HW

LLNL Sierra Center of Excellence Process

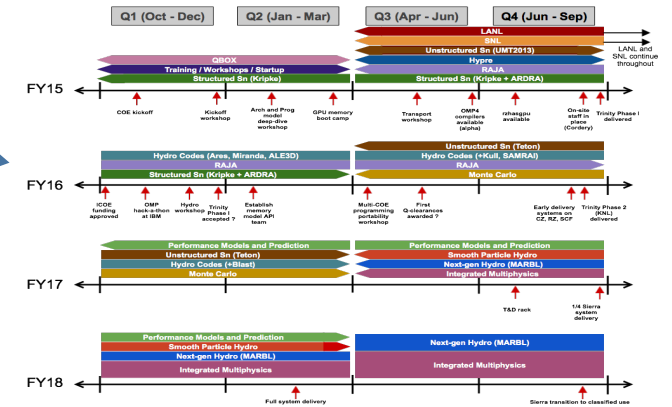
- Work divided into six month work plans
 - 5-6 pages of high level descriptions
 - Deliverables are optimized codes + detailed reports (available upon request)
- 2-4 full-time staff from IBM and NVIDIA
 - Primarily application-focused efforts
 - 1-2 people on site at LLNL, Q-clearances in progress
 - Remote access from IBM Research and NVIDIA
 - Ramps up over time – culminating in big push when machine arrives in 2017-18
- Institutional CoE
 - Adds an additional 2-4 vendor staff
- Coordinated with other NRE work
 - Primarily: Compiler/Tools Working Group

Details captured in a document available upon request:
“Execution Strategy for the LLNL Sierra COE”



Our overall long-term strategy – simplified and summarized

- Codes and libraries “phased in” over time
 - Multi-year plan overlaid with hardware and compiler availability to guide work plans
- Earliest work focused on training and proxy apps
- After year one – pivoted to real applications and greater team engagement with vendor help



(No, you're not supposed to be able to read this!)

Our overarching strategy for each application (1-3 year process)

This is why this meeting is important!

Identify performance *potential* with optimized port (CUDA, alternate algorithms)

- Baseline potential performance

Implement in RAJA, Kokkos, OpenMP 4.x, OpenACC (Fortran)

- Compiler testing
- Establish performance impact

Ensure performance AND portability to other architectures (e.g KNL)

- Performance-portability!!!

**Efforts Documented on an internal Wiki
(requires LLNL authentication, NDA access)**

The screenshot shows the Sierra Center of Excellence Home page. Red annotations highlight key features:

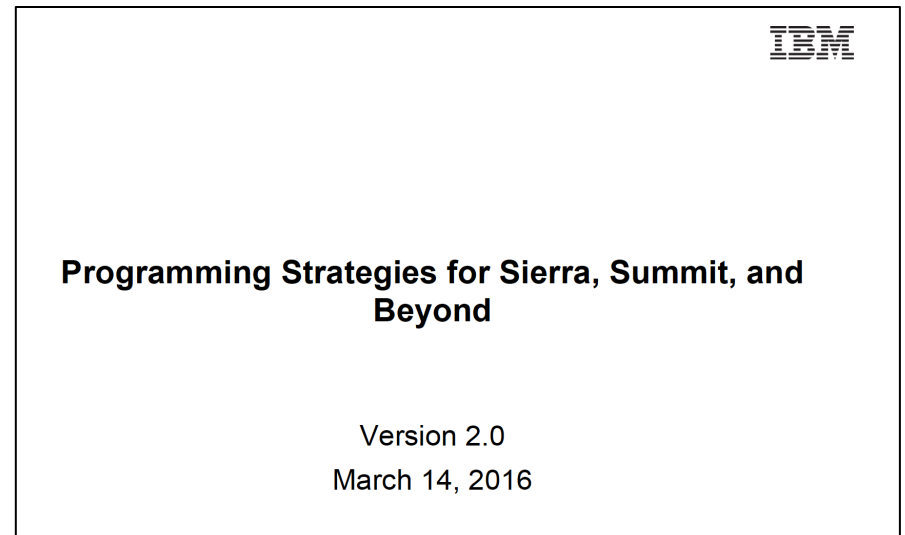
- Quick links:** A row of links at the top: [Sierra Programming Strategies](#), [Platform Plan](#), [ASC Timeline](#), [COE Execution Strategy](#), [ICOE](#), [Trinity COE \(RZ\)](#), and [AAPs](#).
- Past and Upcoming Workshops and CoE Events (most recent first):** A table listing events from April 2016 to December 2014, including DOE Centers of Excellence Performance Portability Meeting, Hydrodynamics themed workshop, and various benchmark deep dives.
- Important Note:** A red box states: "IMPORTANT: You must treat any information on this page that refers to vendor information specific to the Sierra platform as being held under a Non-Disclosure Agreement (NDA). Any discussion of codes that fall under Export Control or UCLN restrictions must not happen on this site. Please use the confluence instance on the Restricted Zone of LG instead." It also advises contacting the page or space author for questions.
- Sierra Logo:** The Sierra Center of Excellence logo, featuring a mountain range and the word "SIERRA".
- Your On-site IBM and NVIDIA Representatives:** A section listing representatives from IBM and NVIDIA, including Matthew Cordery and Steven Rennich, with their contact information.
- Contributing:** A section explaining how to contribute to the site by creating a blog post, with a "Create" button and a "Blog Post" link.
- Currently active labels and linked content:** A section with a table of labels (A-N, O-V, W-Z) and a list of content items (blast, cusparse, fortran, gpu, hydre, icoe, mferm, mps, noh, openmp, questions, scoe-reports, scoenews, sedov, sierra-compilers, software, spmv, tools, workshop).
- News Items:** A list of recent news items, including registration for April Multi-COE Performance Portability Workshop, OpenPower Summit 2016, and new LVM Fortran Solution.
- Recent Blog Posts:** A list of recent blog posts, including registration for April Multi-COE Performance Portability Workshop, OpenMP 4.X Lessons and 5.0 Plans, and Partial OpenMP4.5 GPU support now available.
- Recent space activity:** A list of recent space activity, including Sierra COE Work Plans and Reports, Libraries and Tools project, and CORAL Tools Working Group.
- Callouts:** Red boxes on the right side of the page contain text: "Quick links", "Workshops, training, talks, video, ...", "Contact info for on-site reps", "Subpages by topic breakout, training, papers, CORAL benchmarks, ...", and "Blog posts, news items".

A Sampling of Sierra COE Activities

- Training (e.g. CUDA classes)
 - Accelware offered first course. Planning others (in progress)
- Talks (e.g. invited speakers) and “deep dives”
 - Many captured on video – downloadable from our website
- OpenMP4.x “hack-a-thons”
 - Hands-on experience with early compiler deliveries
 - <https://codesign.llnl.gov/codesign-papers-presentations.php>
- Focused workshops
 - Deterministic transport (May 2015)
 - Hydrodynamics (Dec 2015)
 - Memory hierarchy / NVLINK (planned 2016)
- Targeted application work
 - RAJA, Deterministic transport, Hydrodynamics, Monte Carlo, ...

Joint white paper effort with ORNL *Summit* COE

- *“Programming Strategies for Sierra, Summit, and Beyond”*
- Excellent primer for application teams (~40 pages)
- Living document – updated every 6-12 months
- Volume I:
 - System overview
 - Programming approaches
 - Developing applications
 - Optimization
 - Examples
 - Debugging and Profiling



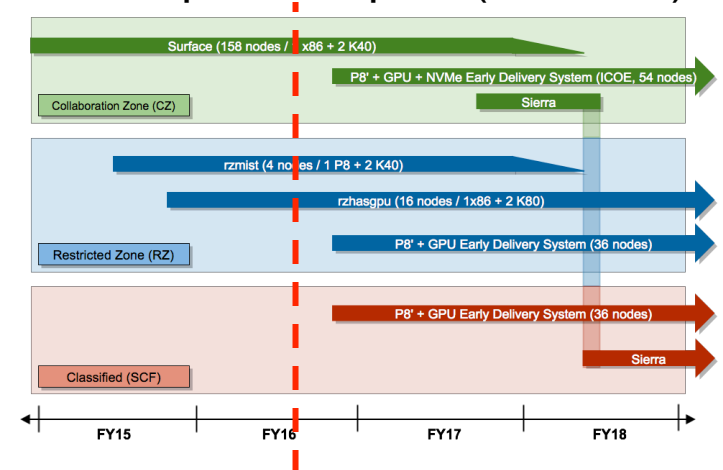
- Volume II in development
 - Will capture specific application examples

Contains **IBM/NVIDIA Confidential**
Available upon request (protect as NDA)

Summary

- 18 months under our belt
 - Halfway to delivery in FY17-18!
- Early delivery hardware anticipated this calendar year
 - 3 systems across different security zones
- LLNL teams beginning to get hands-on access to KNL hardware through Trinity COE

Sierra platform plan (tentative)



- Sierra represents the first ASC has deployed GPUs in a production system
- The COE has been critical to our efforts to get prepared
- We MUST focus on performance AND portability
- Collaborations and teaming are welcomed

